xyzxform

A utility for the manipulation of .xyz files

v2.32

(C) 2005-2009 Jacek Dziedzic

User's Manual

Synopsis:

The **xyzxform** utility is used to manipulate xyz files. An xyz file describes the positions of *N* atoms, termed a *system*. By applying the **xyzxform** utility transformations such as rotations, translations, scaling, selections of subsets of the original system, joining two systems and other may be performed.

The input and output files and the operations to be undertaken are specified in a *directive file*. The **xyzxform** program is then invoked with the name of the directive file as an argument. Consider a simple example – a file named `test.xform` containing the following directives:

```
# A simple directive file
nopbc
zap
in input.xyz
rotate 0 30 0
out output.xyz
```

To execute the operations, the **xyzxform** program is then invoked as follows:

```
xyzxform test.xform
```

Anything following the # character in a directive file, up to the end of line, is considered a comment and is ignored. The first line in the above directive file is an example of a comment. The example directive file realizes a rotation of the system, read from `input.xyz` by 30 degrees about the Y axis. The output is written to `output.xyz`. The directives **nopbc** and **zap** tell **xyzxform** that periodic boundary conditions should not be applied and that the output file should be zapped (erased) before anything is written to it.

Structure of xyz files:

The structure of xyz files is very simple. An xyz file describes the a system of *N* atoms. Each atom is described by a row (line) containing four words – the name of the atom, its *x* coordinate, its *y* coordinate and its *z* coordinate (expressed in Å). For example a gold atom at (8.5, 12.35, -4.2) may be described as

```
Au          8.5000   12.3500  -4.2000
```

The number of spaces between the words and the decimal precision is not significant (here, four decimal digits were used). Scientific notation, where 1.543×10^{-4} would be written as `1.543E-04` may also be used. Thus, the same gold atom could be described with, for instance,

```
Au          8.5E+00 1.235E+01 -4.2E+00
```

The name of the atom is, usually, limited to eight characters. The xyz file contains *N* lines, each describing one of the *N* atoms. The atoms are usually numbered (indexed) from 0 to *N*-1. A two-line header is placed at the very front of the file. The first line contains *N*, the number of atoms. The second line contains a description of the contents of the file and acts like a comment. Several utilities assume that the description begins with a fixed phrase – the word "**BBox:**" followed by six numbers. The numbers are interpreted as the coordinates of two points that are the opposite corners of the simulation box to which the xyz file pertains. For example the following header

```
4000
BBox: 0.0 37.43   0.0 37.43   0.0 37.43 | liquid Cu at 7.97 g/cm3
```

tells us that the xyz file contains 4000 atoms of Cu from a liquid Cu simulation in a box whose corners are at (0, 0, 0) and (37.43, 37.43, 37.43), respectively.

An extension to the xyz format, sometimes termed xyze, is often encountered – it differs only by an addition of a fifth word (often describing the mass or energy of the atom) to the description of each atom.

Several xyz files, describing the same system at different time points, may be concatenated together to form a *multiple-frame* xyz file. Each (2+*N*)-line part of this xyz file is then called a *frame*.

List of directives:

- **input/output:**
  - **in** <xyz_filename>
  - **overwrite**
  - **zap**
  - **out** <xyz_filename>
  - **framerange** <min> <max>
  - **pbc**
  - **nopbc**

- **rotations and translations:**
  - **rotate x** <alpha>
  - **rotate y** <alpha>
  - **rotate z** <alpha>
  - **translate** <x> <y> <z>
  - **center**
  - **periodic_image** <xx> <yy> <zz>

- **scale transformations, changing the size of the simulation box:**
  - **stretch** <xx> <yy> <zz>
  - **rescaleBBox** <xx> <yy> <zz>
  - **BBox** <x1> <x2> <y1> <y2> <z1> <z2>

- **selecting subsets of atoms:**
  - **select cylinder along X** <y> <z> <r>
  - **select cylinder along Y** <z> <x> <r>
  - **select cylinder along Z** <x> <y> <r>
  - **select sphere** <x> <y> <z> <r>
  - **select box** <x1> <x2> <y1> <y2> <z1> <z2>
  - **select above | below plane normal** <a> <b> <c> <d>
  - **select above | below plane through** <x1> <y1> <z1> <x2> <y2> <z2> <x3> <y3> <z3>
  - **select species** <species1> **| any** <species2> **| any** ...
  - **select mass** <m1> <m2> ...
  - **select indices** <imin> <imax>
  - **select all**

- **working with tags:**
  - **tagdef** <tag>
  - **(un)tag cylinder along X** <y> <z> <r>
  - **(un)tag cylinder along Y** <z> <x> <r>
  - **(un)tag cylinder along Z** <x> <y> <r>
  - **(un)tag sphere** <x> <y> <z> <r>
  - **(un)tag box** <x1> <x2> <y1> <y2> <z1> <z2>
  - **(un)tag above | below plane normal** <a> <b> <c> <d>
  - **(un)tag above | below plane through** <x1> <y1> <z1> <x2> <y2> <z2> <x3> <y3> <z3>
  - **(un)tag species** <species1> **| any** <species2> **| any** ...
  - **(un)tag mass** <m1> <m2> ...
  - **(un)tag indices** <imin> <imax>
  - **(un)tag all**

- **other operations:**
  - **lump** <cutoff> <species1> **| any** <species2> **| any** ...
  - **join** <xyz_filename>
  - **uniq | unique** <tolerance>
  - **rename** <species1> **| any** <species2> **| tag | addtag**
  - **name** <new_name>
  - **namestrip**

Overview of directives:

- **input/output:**
  - **`in`** `<xyz_filename>`
    Declares the name of the input xyz file. This directive is mandatory.
  - **`overwrite`**
    Forces the output file to be overwritten when outputting every xyz frame. This is off by default and subsequent frames are appended to the output file, which makes sense if a file containing many frames is transformed. When working with a single-frame input file it might be desirable to have the output file overwritten every time instead of appending to it. This is achieved by using **`overwrite`**.
  - **`zap`**
    A cousin of **`overwrite`**, this directive causes the output file to be zapped (erased) **once**, in the beginning. Subsequent writes to the file will append. Use **`zap`**, and not **`overwrite`** for multiple-frame files.
  - **`out`** `<xyz_filename>`
    Declares the name of the output xyz file. This directive is mandatory. Normally this should be placed as the last directive in the directive file, otherwise the output file may be written out before the transforming directives are applied.
  - **`framerange`** `<min> <max>`
    Forces the program to ignore all the frames whose numbers are not within <`min`, `max`>. This allows you to extract ranges of frames from multiple-frame xyz files. The frames are numbered from 0. Subsequent frames are numbered 1, 2, 3... regardless of the simulation step or time stored in the frame.
  - **`pbc`**
    Turns on periodic boundary conditions. When on, during every operation all the atoms that are outside the simulation box (BBox) will be put back inside the box by the employment of periodic boundary conditions.
  - **`nopbc`**
    Turns off periodic boundary conditions. Atoms remaining outside the simulation box (BBox) will not be put back inside the box.

    Exactly one of **`pbc`** or **`nopbc`** is mandatory.

- **rotations and translations:**
  - **`rotate x`** `<alpha>`
    Rotates the system about the X axis by an angle `alpha`, expressed in degrees. BBox is not altered. Rotating about an axis *parallel* to the X axis is possible by a combination of rotation and translation.
  - **`rotate y`** `<alpha>`
    Rotates the system about the Y axis by an angle `alpha`, expressed in degrees. BBox is not altered. Rotating about an axis *parallel* to the Y axis is possible by a combination of rotation and translation.
  - **`rotate z`** `<alpha>`
    Rotates the system about the Z axis by an angle `alpha`, expressed in degrees. BBox is not altered. Rotating about an axis *parallel* to the Z axis is possible by a combination of rotation and translation.

  - **`translate`** `<x> <y> <z>`
    Translates the system (all atoms) by a vector [`x`, `y`, `z`]. BBox is not altered.
  - **`center`**
    Translates the system (all atoms) in such a way, that the centre of mass is at (0, 0, 0). NB: The current version assumes that all atoms have the same mass. BBox is not altered.
  - **`periodic_image`** `<xx> <yy> <zz>`
    Replaces the system (all atoms) with its periodic image, by translating all atoms by a vector that is a [`xx`, `yy`, `zz`]-tiple of the box. BBox is not altered. A typical usage would be to subsequently rename the atoms and to `join` with the original file, thus obtaining the original system plus its desired periodic image.

- **scale transformations, changing the size of the simulation box:**
  - **`stretch`** `<xx> <yy> <zz>`
    Scales (stretches) the system by a factor of `xx` along the X axis, a factor of `yy` along the Y axis, a factor of `zz` along the Z axis. BBox is not altered. Scaling is realized by multiplying all the coordinates by the given factor. Because of this care must be taken if the BBox does not begin at (0, 0, 0).
  - **`rescaleBBox`** `<xx> <yy> <zz>`
    Scales (stretches) the BBox by a factor of `xx` along the X axis, a factor of `yy` along the Y axis, a factor of `zz` along the Z axis. The atomic positions remain unchanged. Typically used alongside **`stretch`**.
  - **`BBox`** `<x1> <x2> <y1> <y2> <z1> <z2>`
    Sets up a new BBox. Atomic positions remain unchanged.

- **selecting subsets of atoms:**
  Warning: All the selection operators (performed by the **`select`** directive) described above always replace the system with the new one, which contains only the selected atoms. Because of this, subsequent **`select`** operations will always result in selecting smaller and smaller subsets of the original system (the intersection of all the selections). If you need to obtain a sum of several subsets, you will be better off using atom tagging (the **`tag`** directive).

  The **`select cylinder`** directives allow for a selection of atoms contained within a cylinder parallel to one of the cartesian axes X, Y or Z. If you need to select atoms within a differently oriented cylinder, proceed as follows: Turn off periodic boundary conditions, rotate the system so that the desired cylinder is parallel to one of the cartesian directions, select the atoms within the cylinder and rotate the system back.

  - **`select cylinder along X`** `<y> <z> <r>`
    Selects atoms lying within an infinintely long cylinder or radius `r`, parallel to the X axis, whose axis lies on (-∞, `y`, `z`) to (∞, `y`, `z`).
  - **`select cylinder along Y`** `<z> <x> <r>`
    Selects atoms lying within an infinintely long cylinder or radius `r`, parallel to the Y axis, whose axis lies on (`x`,-∞, `z`) to (`x`,∞, `z`).
  - **`select cylinder along Z`** `<x> <y> <r>`
    Selects atoms lying within an infinintely long cylinder or radius `r`, parallel to the Z axis, whose axis lies on (`x`,`y`,-∞) to (`x`,`y`,∞).

  - **`select sphere`** `<x> <y> <z> <r>`
    Selects atoms lying within a sphere of radius `r`, and a centre at (`x`,`y`,`z`).
  - **`select box`** `<x1> <x2> <y1> <y2> <z1> <z2>`
    Selects atoms lying within a cuboid whose opposite corners are at (`x1`, `y1`, `z1`) and (`x2`, `y2`, `z2`).
  - **`select above plane normal`** `<a> <b> <c> <d>`
    Selects atoms lying above a plane given by the plane equation `a`x+`b`y+`c`z+`d`=0, that is atoms whose coordinates satisfy `a`x+`b`y+`c`z+`d`>0 are selected. Geometry reminder: The vector [`a`, `b`, `c`] is normal to the plane and `d` controls the offset of the plane along this vector. For instance, to select atoms lying above a plane parallel to the Z axis, one needs to take `c`=0, and then choose `a`/`b`=tg(θ), where θ is the angle between the plane of interest and the ZX plane. NB: Depending of what is desired, it may be easier to use **`select above plane through`**.
  - **`select below plane normal`** `<a> <b> <c> <d>`
    As above, only atoms whose coordinates satisfy `a`x+`b`y+`c`z+`d`<0 are selected.

  - **`select species`** `<name1> <name2> ...`
    Selects atoms of the species (name) supplied. The names must match exactly. The special name **`any`** may be used to match every possible name. Whereas this is pointless for the **`select`** directive, this may be used with the **`tag`** and **`untag`** directives.

  - **`select mass`** `<m1> <m2> ...`
    Selects atoms with the specified masses, within an absolute tolerance of 1E-10. Of course only .xyze files specify atomic masses.

  - **`select indices`** `<imin> <imax>`
    Selects atoms with the specified indices (numbers), that is atoms whose numbers lie within `<imin`,`imax>`. Atoms are numbered starting from 0.

  - **`select all`**
    Selects all atoms, thus doing nothing. Whereas this makes little sense with the **`select`** directive, it might be useful to **`tag all`** or **`untag all`** (see **working with tags**).

- **working with tags:**

  Basic usage. All the **tag** directives *tag* certain atoms, those which fulfil a specified condition and the **untag** directives *untag* specified atoms. The conditions are the same as for the **select** directive described above. In contrast to the application of **select**, the system is **not** replaced by a new one – the atoms that are not tagged still exist, they are however not written to the output file when the **out** directive is encountered. Initially all atoms are tagged, thus you would normally first **untag** a subset of atoms or all the atoms (**untag all**). By combining multiple **tag** and **untag** directives complex subsets of the system may be built. Once again, it is only the atoms that are tagged that are written out to the output file.

  Advanced usage. Atoms may be tagged with differing tags. The tag is specified as a single character, the default tag being "&", the current tag is specified by the **tagdef** directive. Tagging is, in general, transparent, that is the character used to tag an atom does not influence its name, location or mass. The tag itself is also not written out to the output xyz file – the only way to reveal the tag is to use the **rename** <name> tag directive, which will rename the atom to its tag or to use the **rename** <name> addtag, directive, which will concatenate the tag to the name of the atom.

  Example:

```
# All atoms within a certain cylinder are renamed by
# adding ^ to their name. Other atoms are unaffected.
nopbc
zap
in in.xyz
untag all
tagdef ^
tag cylinder along X 0 0 8
rename any addtag
tag all           # so that all atoms are written out
out out.xyz
```

- **other operations:**
  - **`lump`** `<cutoff> <species1>` **`| any`** `<species2>` **`| any ...`**
    An advanced operation. Classifies the atoms into classes basing on their connectivity. For atoms of the given species (or for all atoms if **`any`** was specified) the following procedure is performed. The atom with the lowest index is chosen and its mass is set to 1 (which will be termed *the atom belongs to class* 1). Then all the neighbours of this atom (that is, all atoms no further than `cutoff` from the atom) are also placed into class 1. The procedure is repeated for the newly added atoms (that is, the neighbours of the neighbours, their neighbours and so on, are all placed into class 1). As a consequence, all atoms which may be reached by starting from the initial 'seed' atom by hopping through other atoms, no further than `cutoff` each time. If any atoms remain, the one with the lowest index is chosen and it is placed into class 2, then the procedure is repeated for class 2. The classification procedure is repeated, with increasing class numbers, until all the atoms have been classified.

    As a result all the atoms are placed into classes – atoms which are members of class *N* all belong to a cluster (in the sense of neighbourhood dictated by `cutoff`). The classes may be inspected by looking at the masses in the output xyze file. If you desire to change the *names* of the atoms to reflect their class membership, a combination of **`tagdef`**, **`tag mass`**, **`rename`** `<name> tag` (or **`rename`** `<name> addtag`) may be used.

  - **`join`** `<xyz_filename>`
    Merges the atoms in a xyz file `xyz_filename` into the system, placing them at the bottom (indexwise). BBox is not altered.

  - **`uniq | unique`** `<tolerance>`
    For all atoms that occupy the same point in space (up to a `tolerance` along each cartesian direction) only the last one is retained. In this way duplicates, e.g. resulting from joining several files, can be eliminated.

  - **`rename`** `<species1>` **`| any`** `<species2>` **`| tag | addtag`**
    Renames all atoms with a name of `species1` to `species2`. The special name **`any`** may be used instead of the source name to rename all atoms. The special name **`tag`** may be used instead of the target name to change the name of the atoms to their *tag character* (cf. **working with tags**). The special name **`addtag`** may be used instead of the target name to rename the atoms by the concatenation of their *tag character* to their current name (cf. **working with tags**).

  - **`name`** `<new_name>`
    Changes the name of the system (the string in the first line of the frame, placed after the BBox).
  - **`namestrip`**
    By default, the declared name of the system is appended to the previous name, along with the previous BBox, to preserve a trace of the changes to the system. If this is not desired and you want the old names and BBoxes to be left out, specify **`namestrip`** .

Other concerns:

The comment char # may be at any time intepreted literally if it is escaped by a backslash. Thus, the sequence \# is interpreted as a non-comment # character.

Examples of directive files:

```
# Selects atoms within a sphere with a radius of 12 Å, centered at (-4,3,8.5)
# Periodic boundary conditions are not applied
nopbc
zap
in input.xyz
select sphere -4 3 8.5 12
out output.xyz
```

```
# Extracts only frames 10 to 19 from the input file f.xyz
# Periodic boundary conditions are not applied
nopbc
zap
in f.xyz
framerange 10 19
out output.xyz
```

```
# Renames all the 'Au' atoms to 'Pt'
nopbc
zap
in input.xyz
rename Au Pt
out output.xyz
```

```
# Renames all the Au atoms within a box (0,0,0)-(10,10,10) to 'B'
# Periodic boundary conditions are in effect
pbc
zap
in input.xyz
untag all
tagdef B
tag box 0 10 0 10 0 10
rename Au addtag
tag all
out output.xyz
```

```
# Stretches the system by 10% along the Z direction
# Periodic boundary conditions are not applied
nopbc
zap
in input.xyz
stretch 1 1 1.1
rescaleBBox 1 1 1.1
out output.xyz
```